Cluster Computing Development and Research

PI: Paul von Allmen, JPL

Task Completion Report

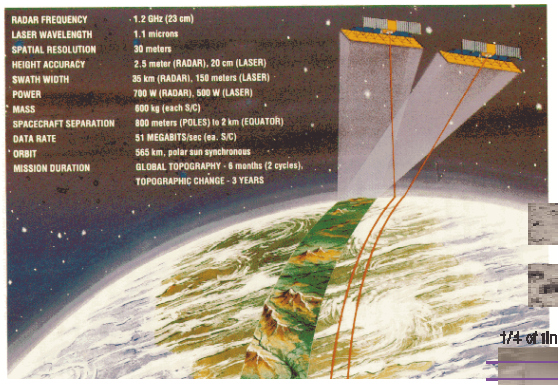
Computational Technologies - **Cluster Computing**

Cluster Computing Development and Research

PI: Paul von Allmen, JPL

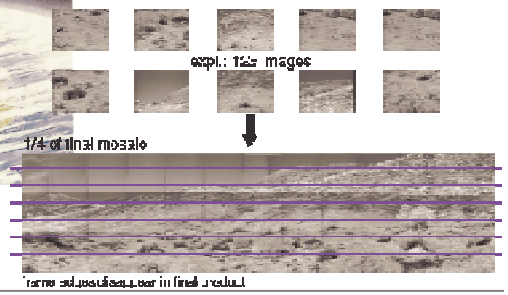
Objective

- Examine cluster computers for I/O limitations
- Compare cluster I/O to supercomputers for cost effectiveness
- Use scientific data processing application as test cases:
  - Parallel mosaic processing of Mars images.
  - Repeat Orbit Interferometry data processing (ROI_PAC)



Repeat Orbit Interferometry

Mars Image Mosaic



Approach

- Implement state-of-the-art networking hardware.
- Implement RAID (redundant array of inexpensive disks) system for parallel I/O.
- Parallelize scientific data processing software using Message Passing Interface
- Measure application performance, including I/O, on both clusters and SGI supercomputer

Partner: P.Rosen,JPL

Key Milestones

• Install and test Myricom hardware	5/02
• Study gigabyte networking systems	9/02
• Mars Image Mosaicing on RAID system	1/03
• Optimize ROI_PAC on cluster with RAID	5/03
• Improved SRTM data processor on cluster	10/03
• Report on parallel I/O work	11/03
• Include parallel computation and I/O technology in Open Source distribution of ROI_PAC data processing software.	6/04

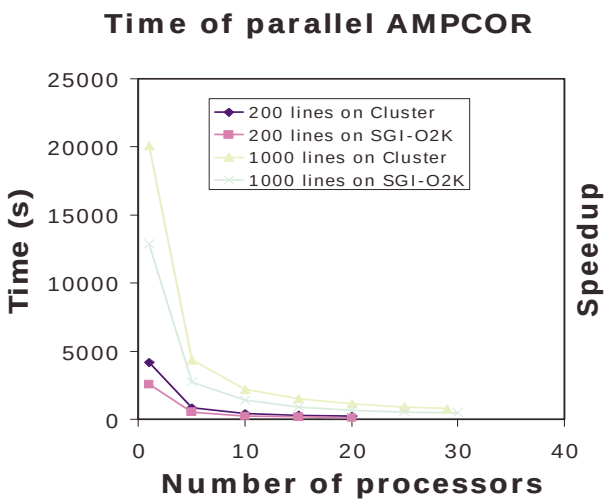
TRL_{in} =4

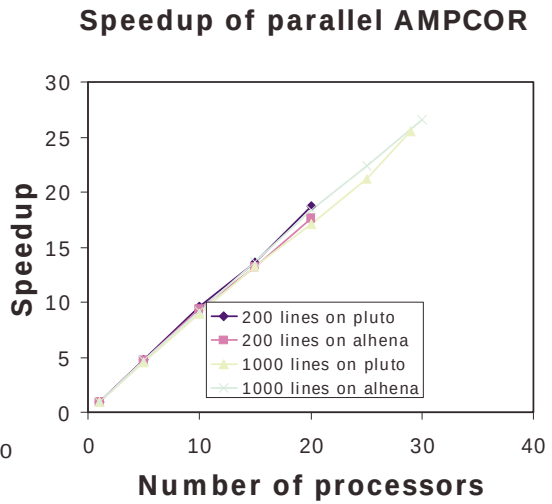
Cluster Computing Development and Research

PI: Dr Paul von Allmen at JPL

Objective

- Examine cluster computers for I/O limitations
- Compare cluster I/O to supercomputers for cost effectiveness
- Use scientific data processing application as test cases:
  - Parallel mosaic processing of Mars images.
  - Repeat Orbit Interferometry data processing (ROI_PAC)





Accomplishments

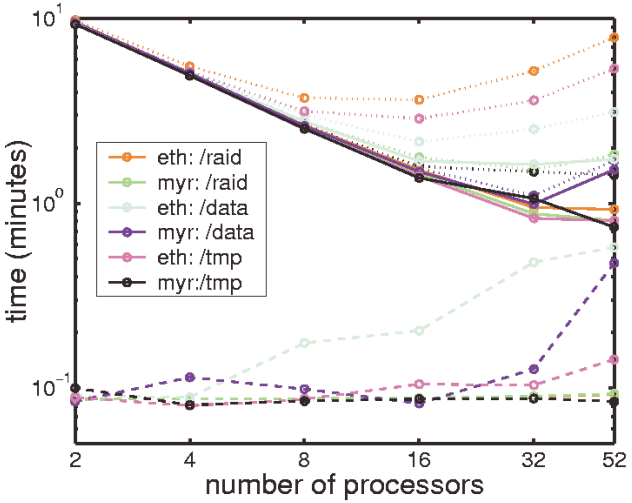
- The Parallel Virtual File System was used on cluster computers to overcome I/O issues
- Parallelized computational and I/O tasks in two subroutines of ROI_PAC (Repeat Orbit Interferometry Package for Synthetic Aperture Radar software) which constitutes most of the runtime.
- Performance on the cluster computer was comparable to performance on SGI supercomputer, at 1/10 the hardware cost.
- Sets framework for efficient parallelization of computational and input/output tasks on cluster computers and matches performance on costly shared memory architecture.
- Achieves faster turn around time for scientific data processing software at a lower cost.
- Parallel extensions will be included in the next public release of the Open Source ROI_PAC, available on OpenChannel Foundation web site.

Partner: P.Rosen/JPL

TRL_{in} =4; TRL_{out}=5

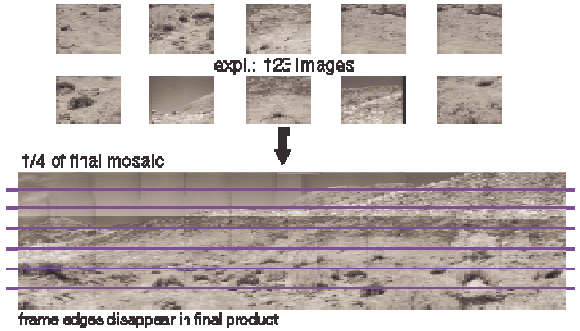
Cluster Computing Development and Research

Task Leader: Paul von Allmen, JPL

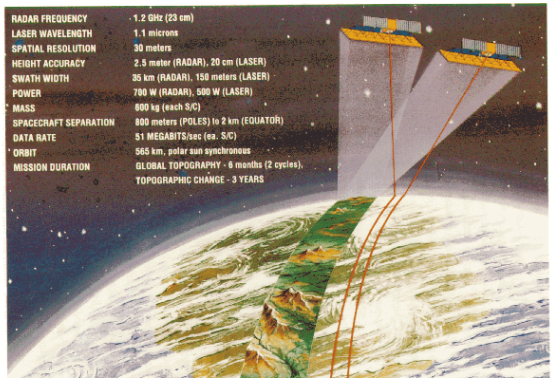


RAID (Redundant Array of Inexpensive Disks) I/O System on Parallel Mosaic Software:

- **Without RAID, I/O time increases with number of CPU (reaches 30% of execution time with 52 CPU).**
- **With RAID, I/O time stays constant with number of CPU**



Mars Image Mosaic



Repeat Orbit Interferometry ROIPAC

Optimize Performance on Cluster:

- **Implement coarse-grain parallelization over data patches in AMPCOR.**
- **Speed-up on SGI O2K is 25 on 30 CPUs.**
- **Speed-up on cluster is 25 on 30 CPUs.**

Time of parallel AMPCOR